IEEE Big Data, 2021

HAR: Hardness Aware Re-weighting for Imbalanced Datasets





Rahul Duggal
Georgia Tech

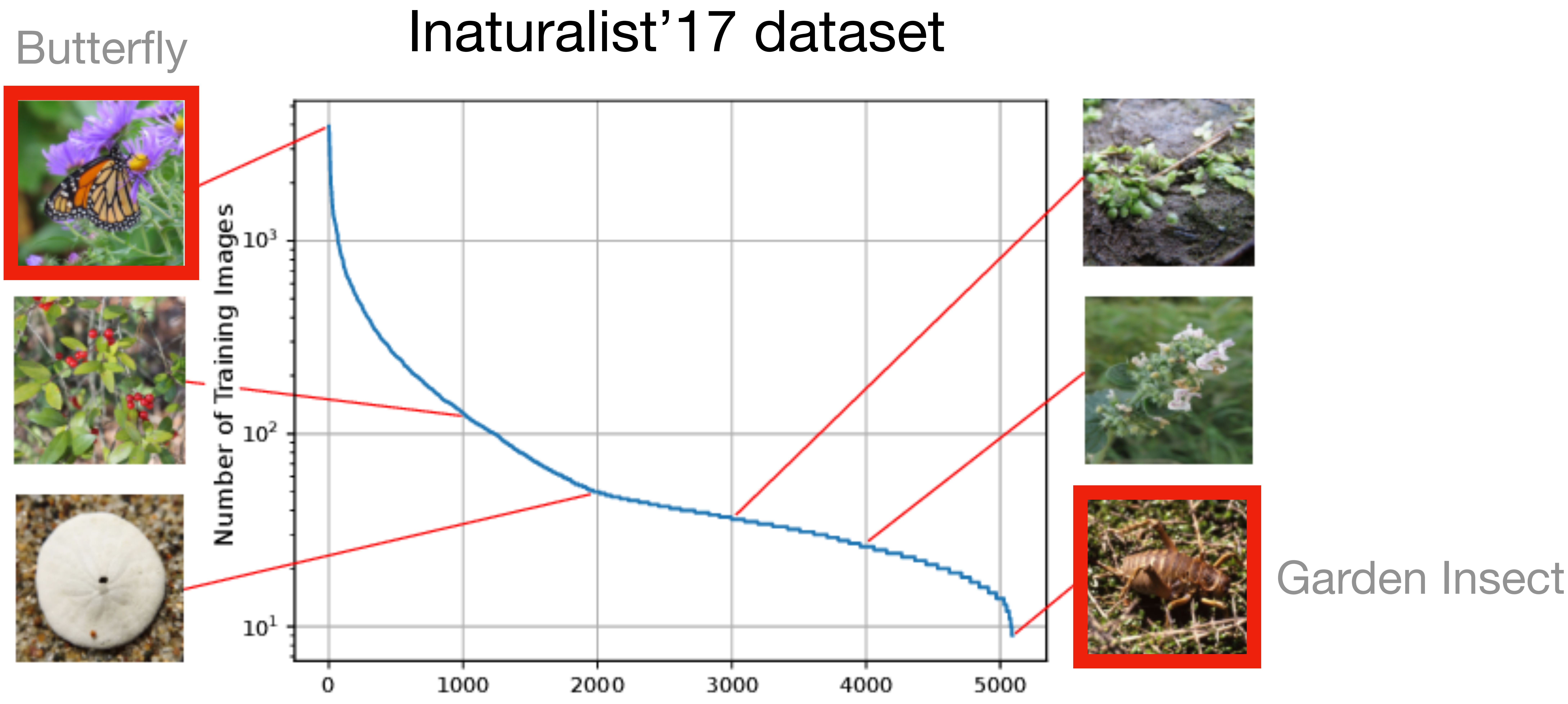
Scott Freitas
Georgia Tech

Sunny Dhamnani
Georgia Tech

Polo Chau
Georgia Tech

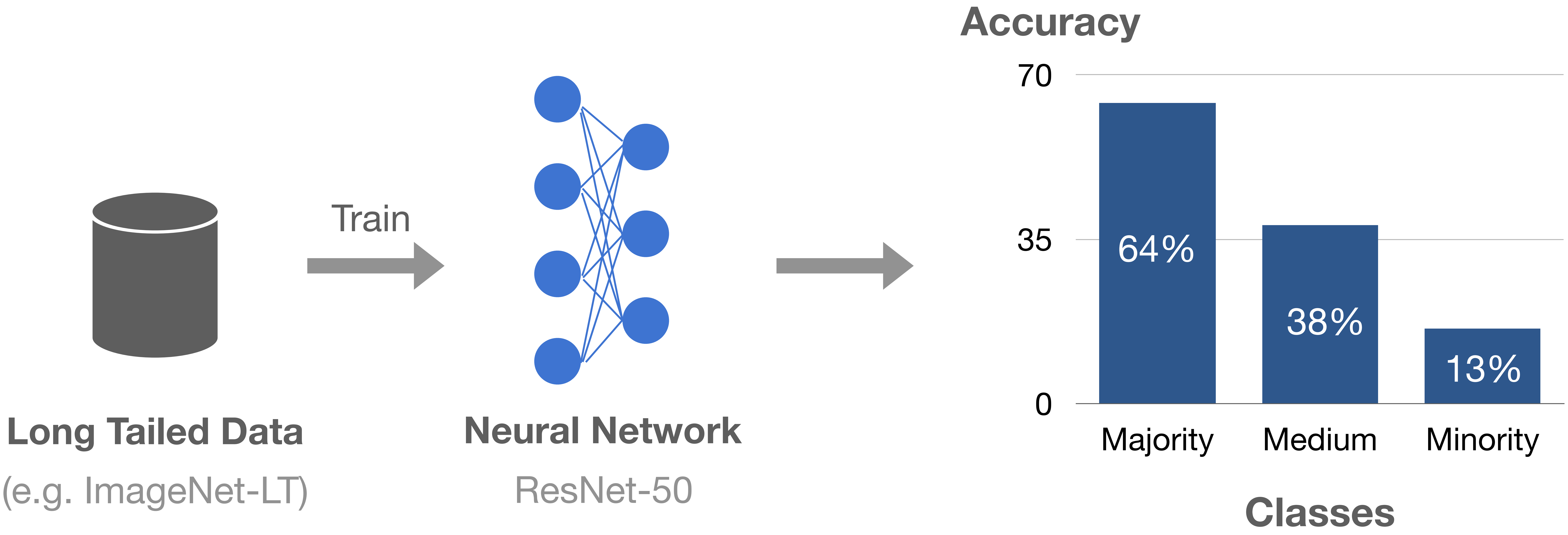
Jimeng Sun
UIUC

Challenging long tail of data



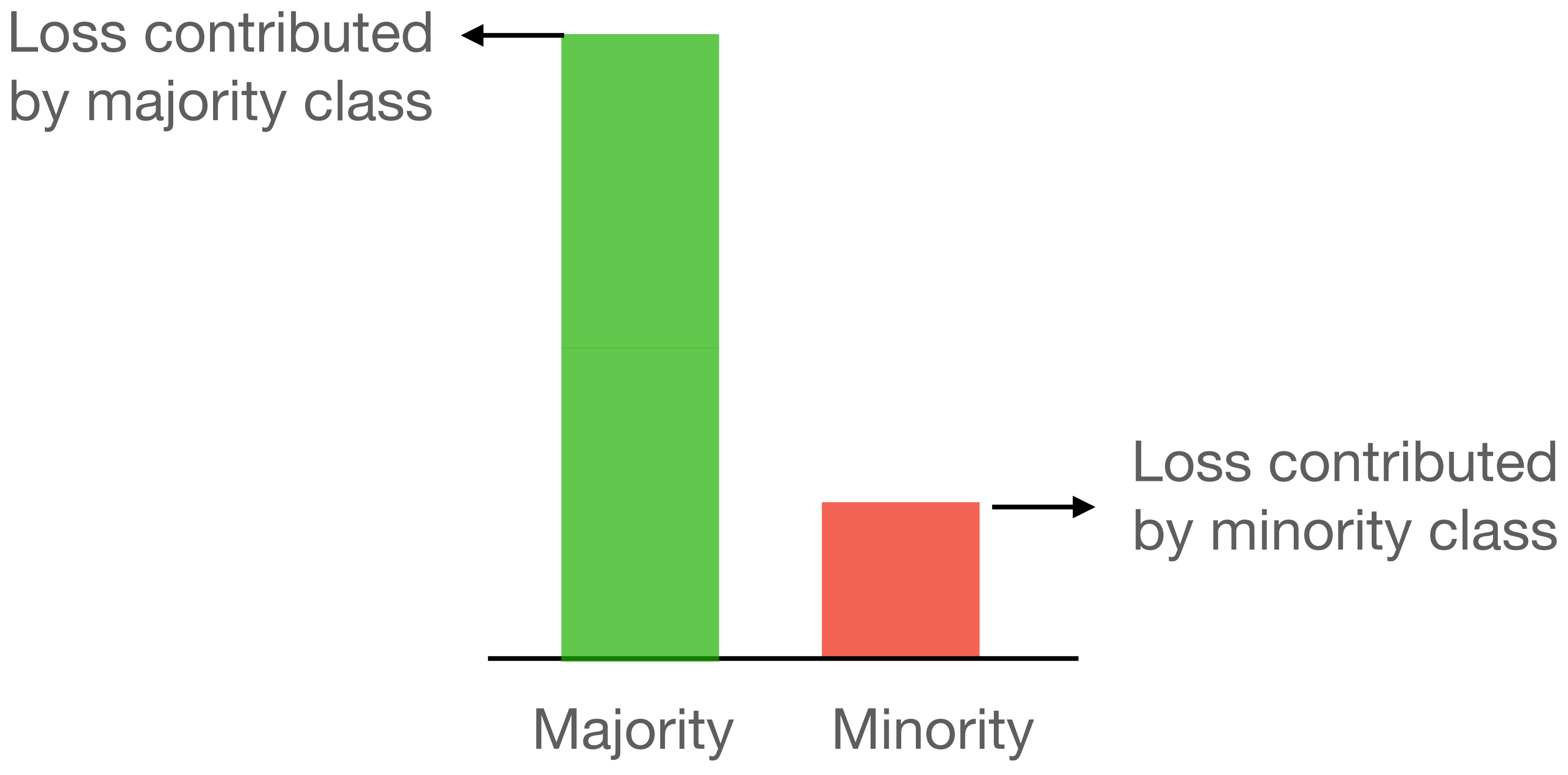
Horn, Grant Van et al. "The iNaturalist Species Classification and Detection Dataset." 2018 IEEE/CVF Conference on Computer Vision and Pattern Recognition (2018): 8769-8778.

Challenging long tail of data

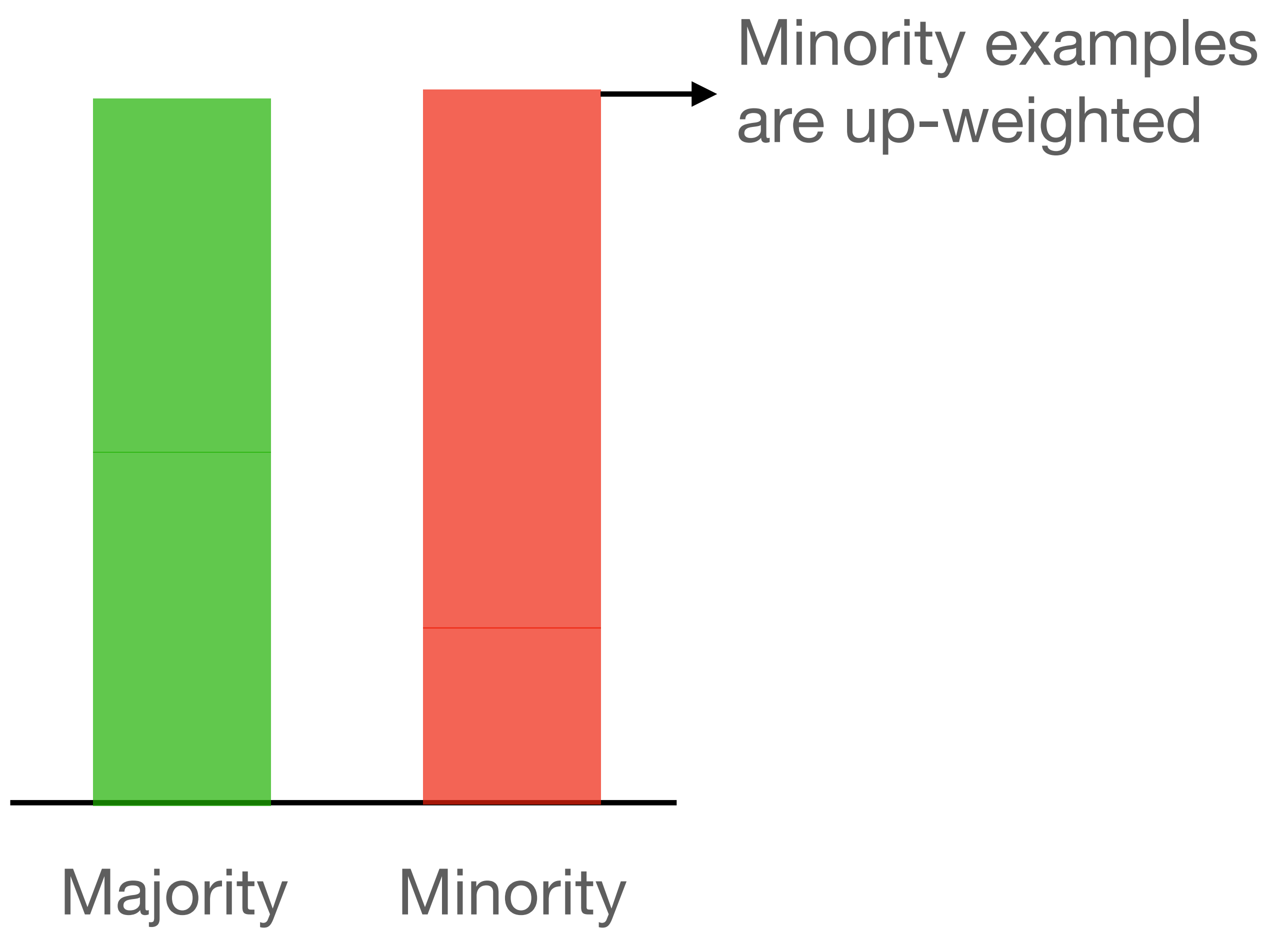


Loss Re-weighting

Loss Re-weighting



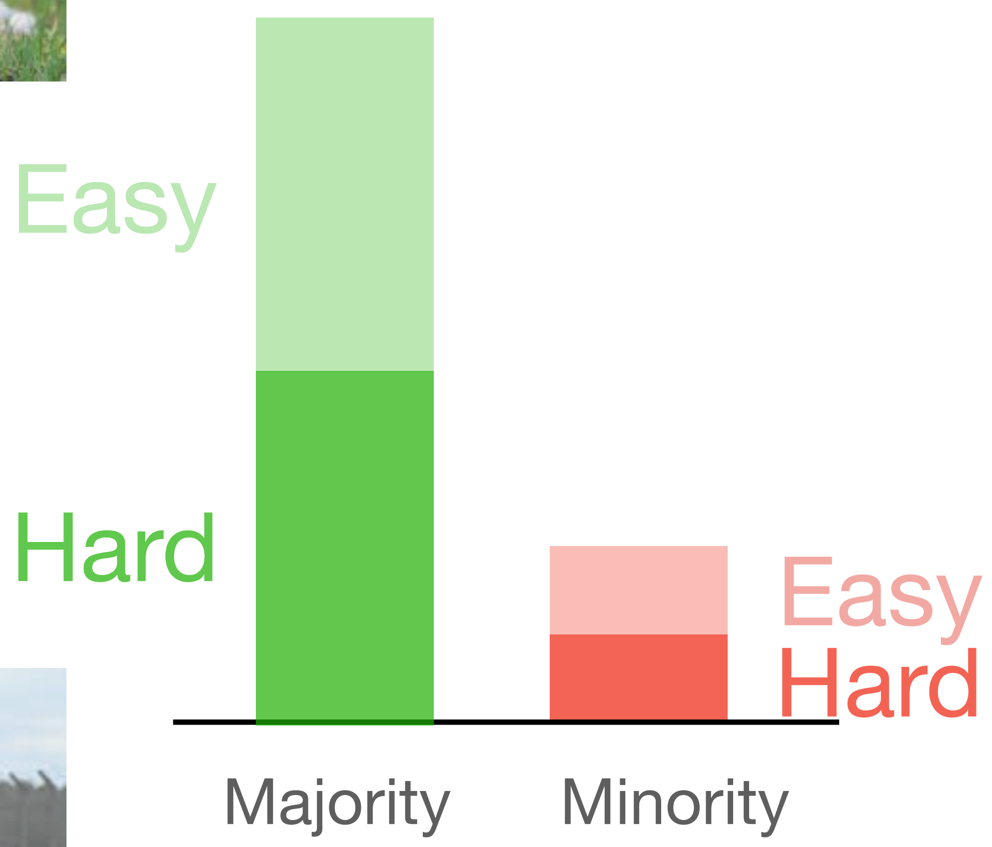
Original



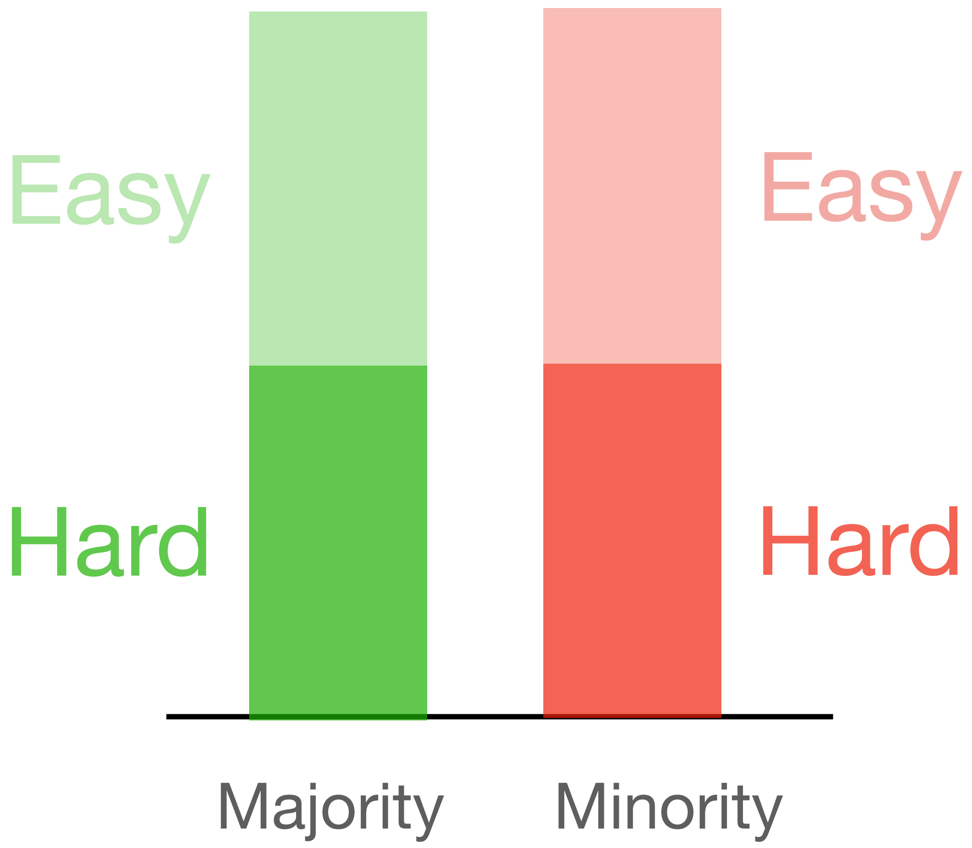
Uniform Re-weighting

Loss Re-weighting

Loss Re-weighting



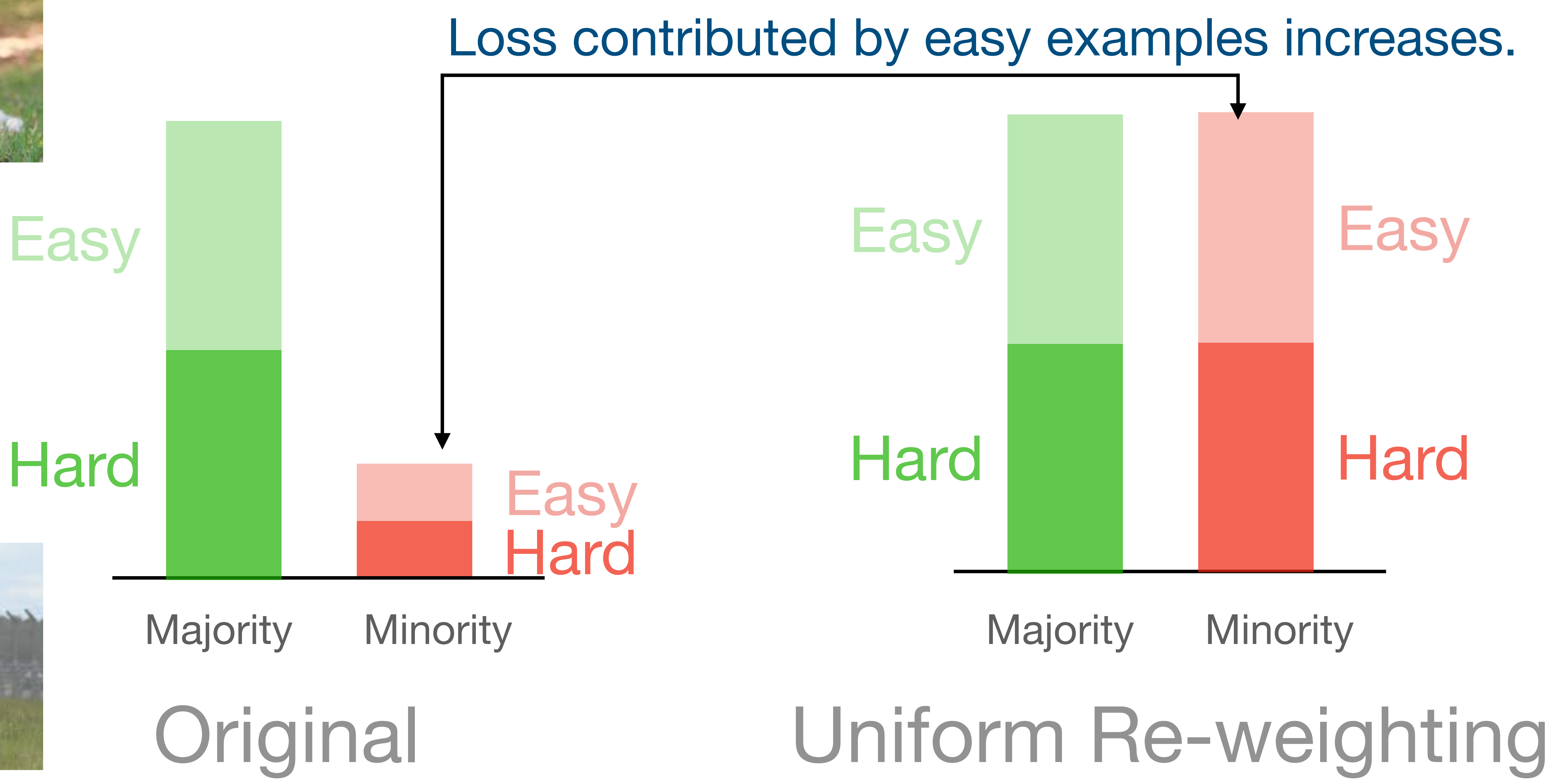
Original



Uniform Re-weighting

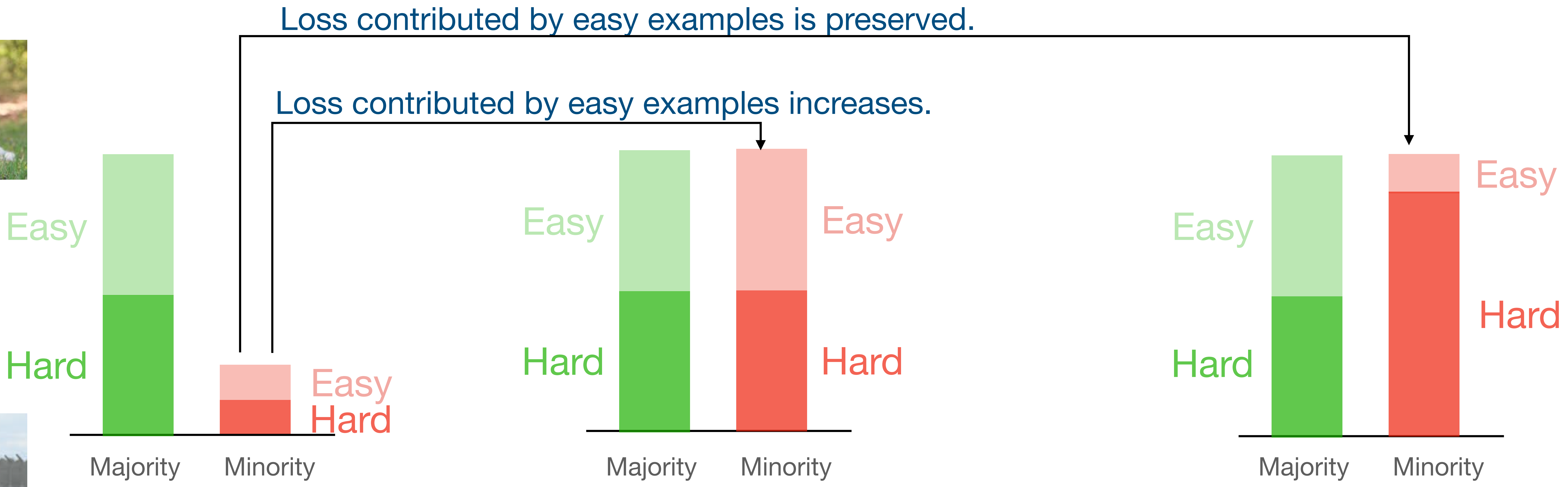
Easy and hard examples
up-weighted **equally**

Loss Re-weighting



Easy and hard examples
up-weighted **equally**

Loss Re-weighting



Original

Uniform Re-weighting

Hardness Aware Reweighting
(ours)

Easy and hard examples
up-weighted **equally**

Hard examples up-weighted
more than easy examples

Loss Re-weighting

Hardness Aware Reweighing (HAR) strategy that up-weights only the **hard examples!**

Challenge

No labels for hardness!

Dogs

Easy?

Hard?

Cats

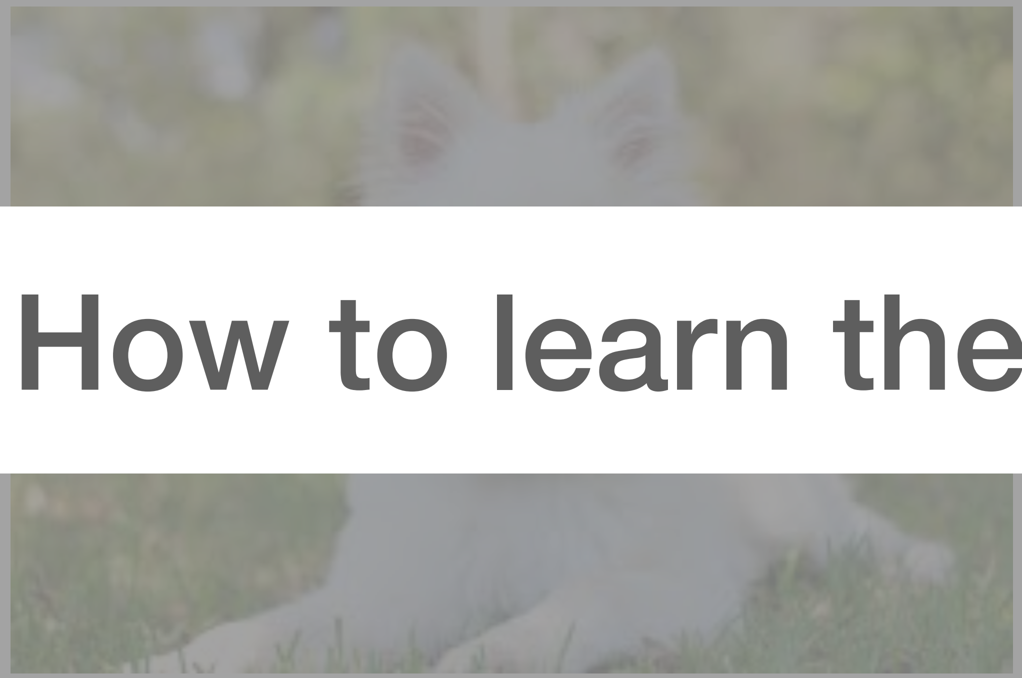
Easy?

Hard?

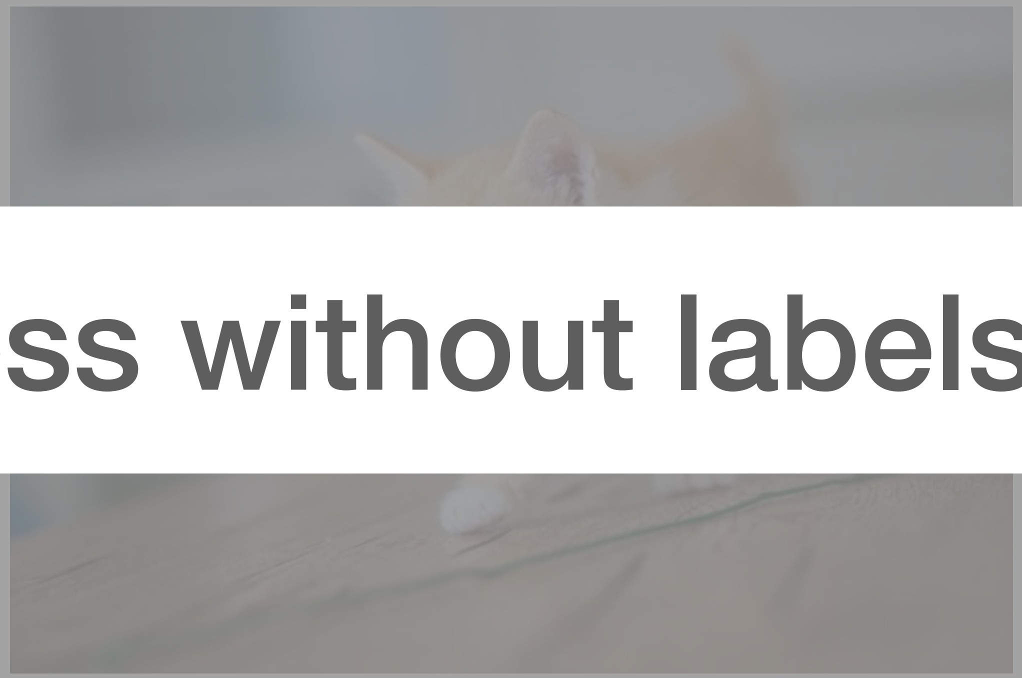
Challenge

No labels for hardness!

Dogs



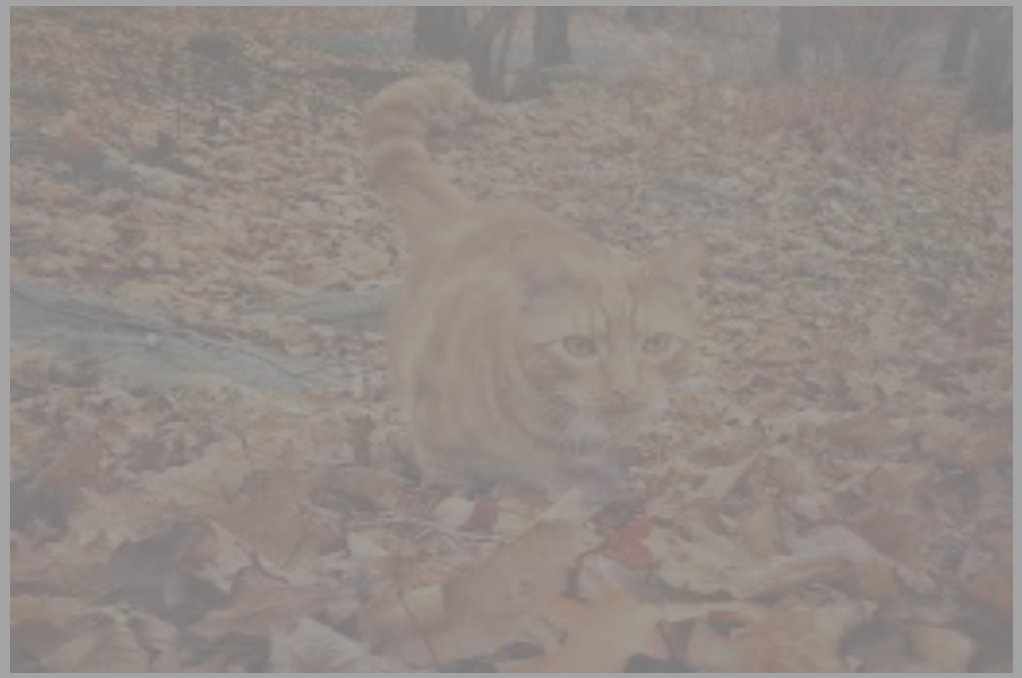
Cats



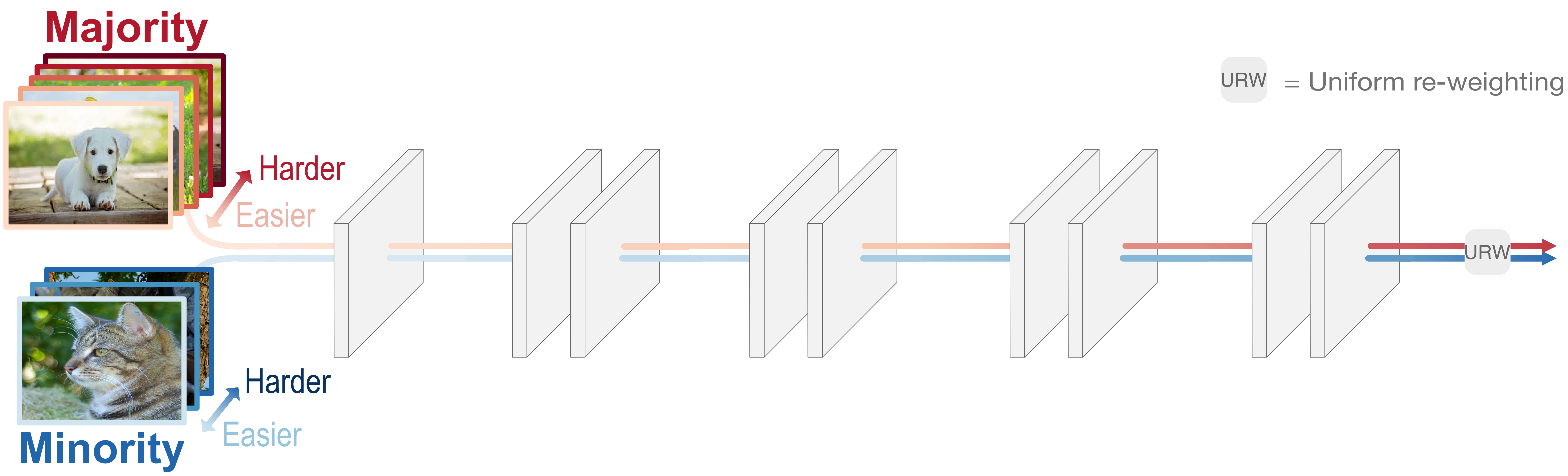
How to learn the notion of hardness without labels?

Hard?

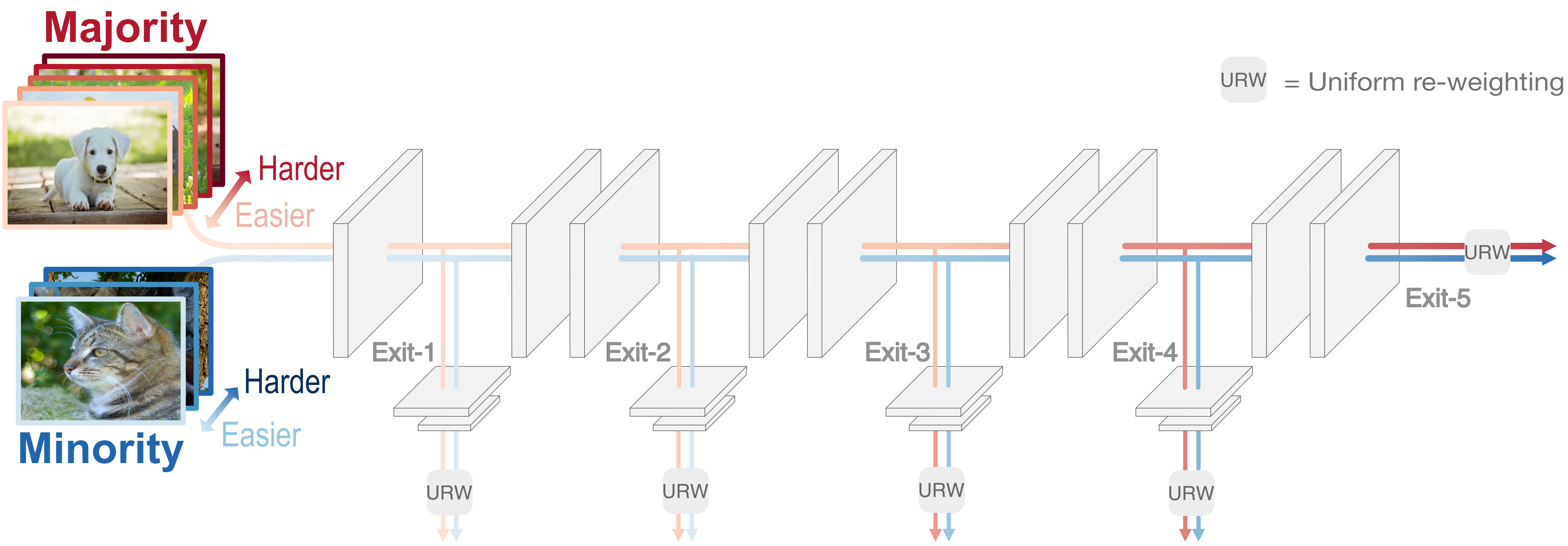
Hard?



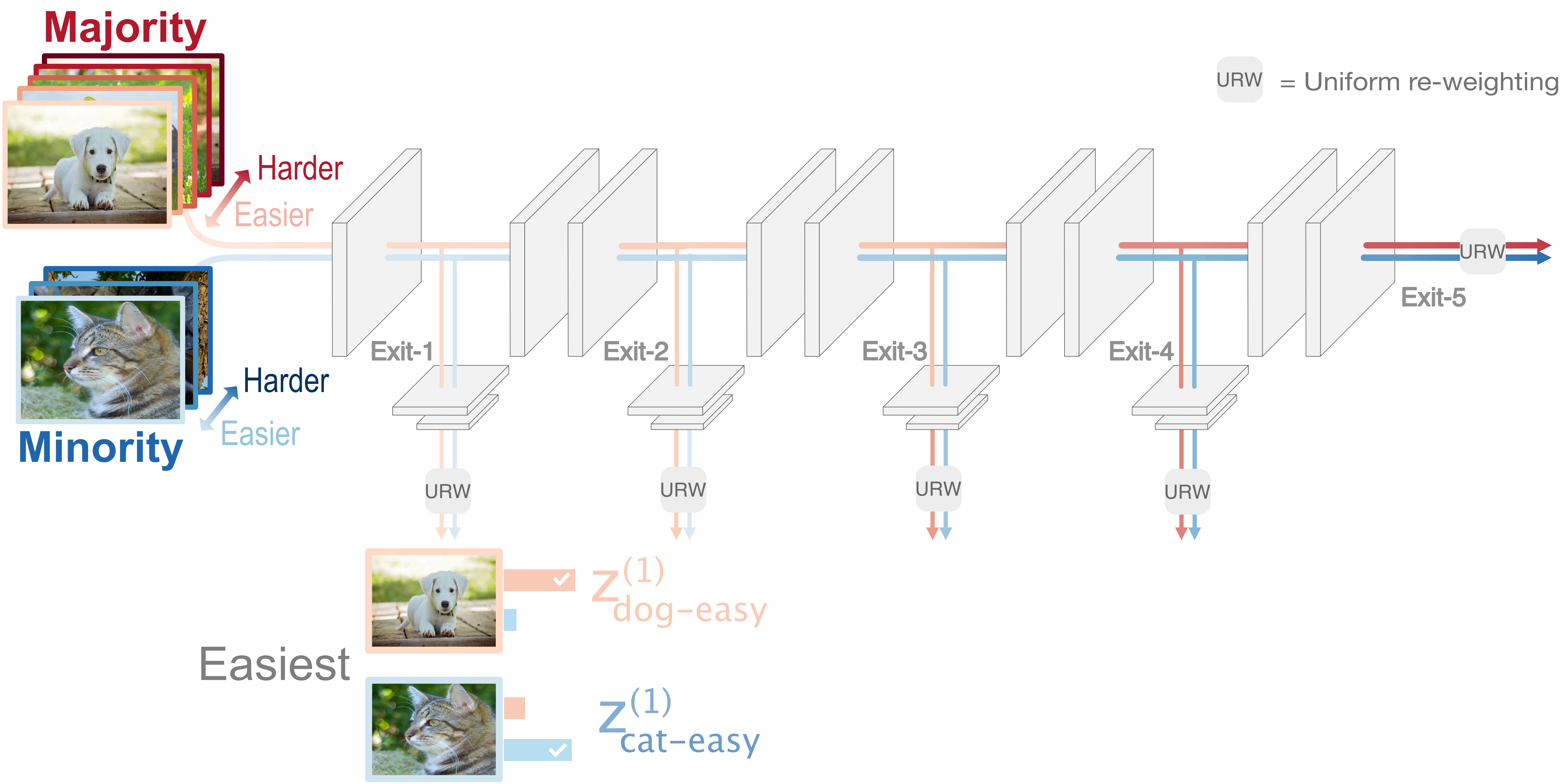
HAR Framework



HAR Framework



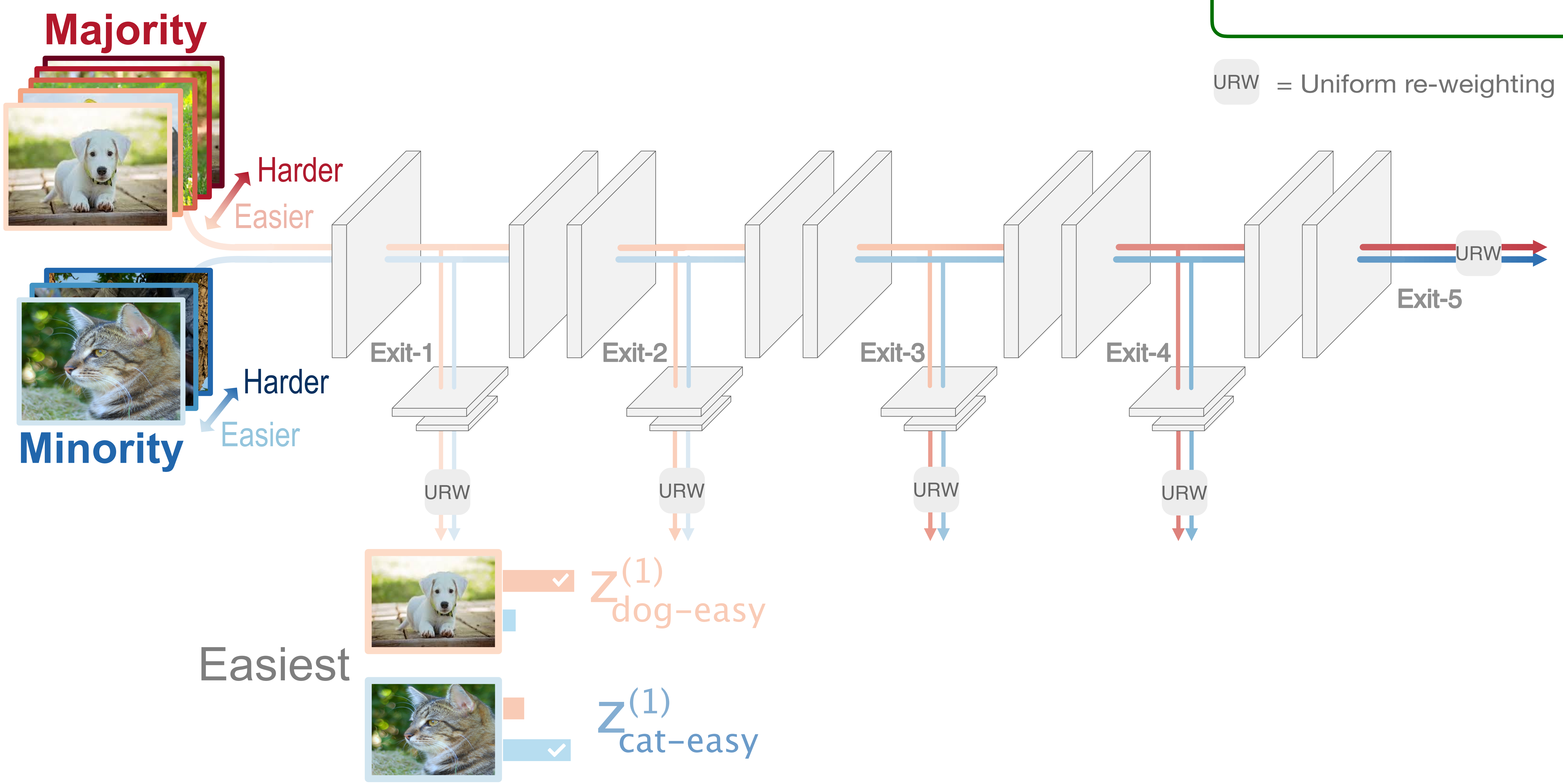
HAR Framework



HAR Framework

Early-Exit Conditions for Image

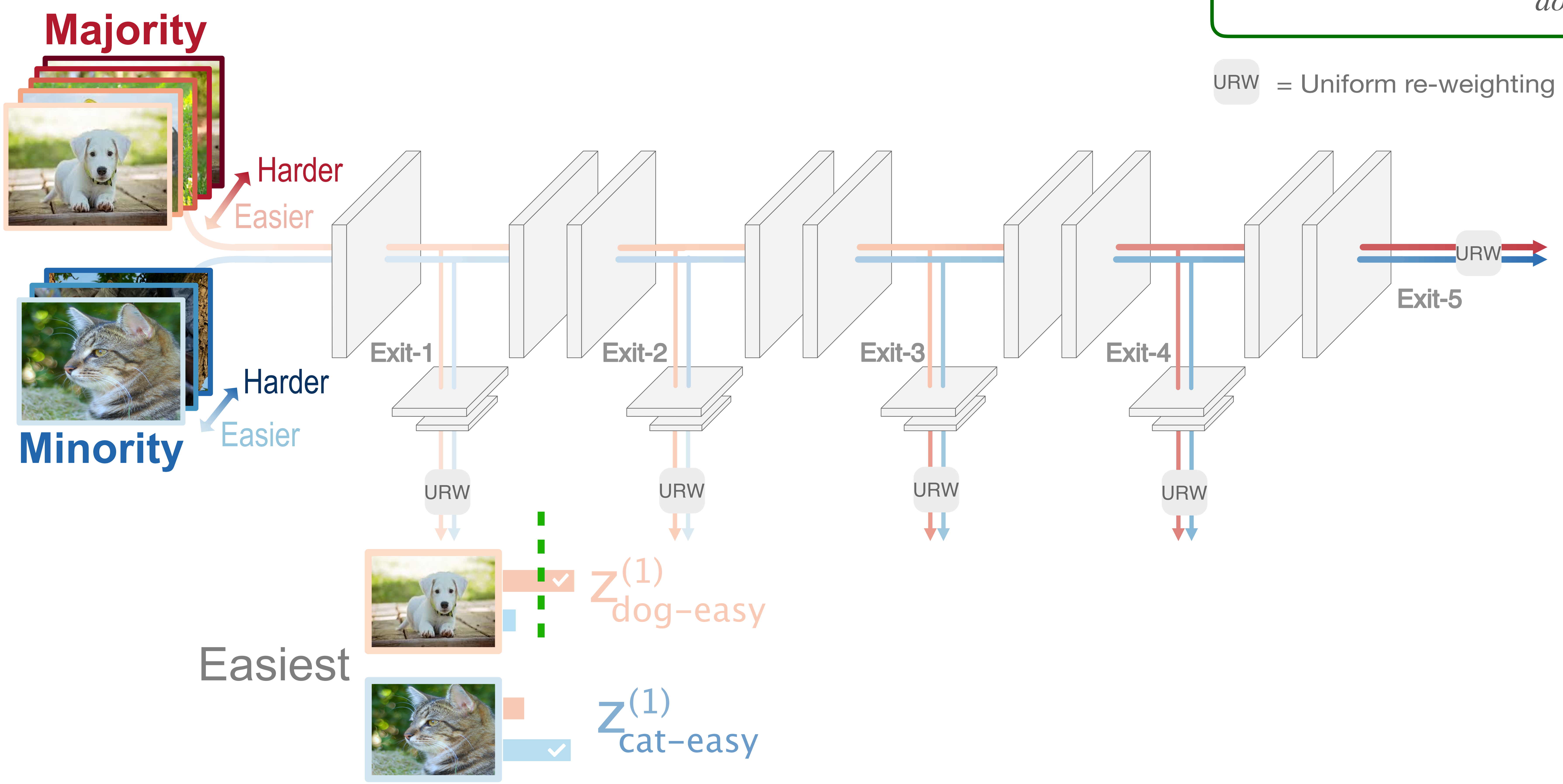
1. Correct $argmax(z^{(1)}_{dog-easy}) = y_{dog}$



HAR Framework

Early-Exit Conditions for Image

1. Correct $argmax(z^{(1)}_{dog-easy}) = y_{dog}$
2. Confident $z^{(1)}_{dog-easy}[dog] > t$

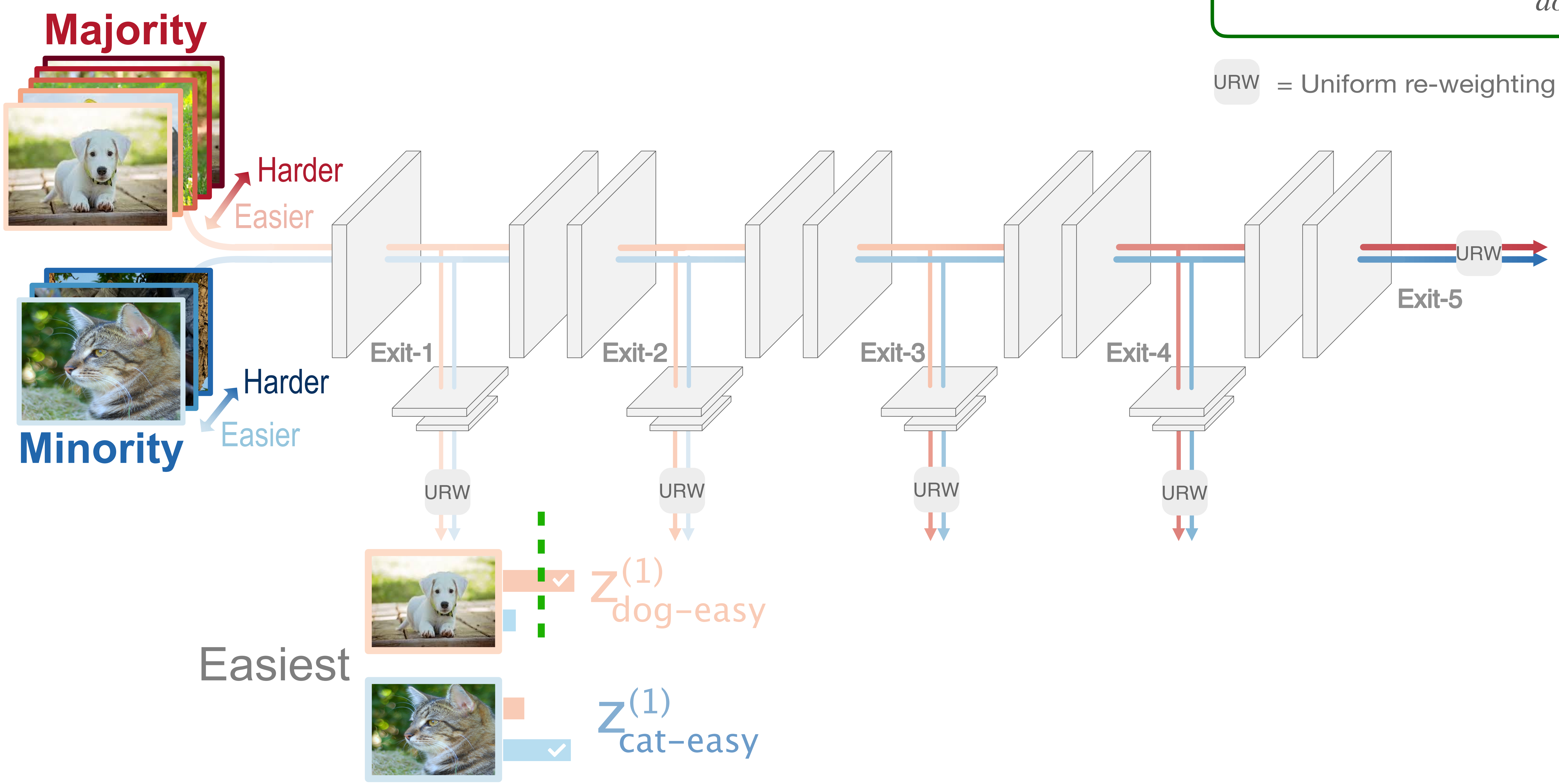


HAR Framework

Early-Exit Conditions for Image

1. Correct $argmax(z^{(1)}_{dog-easy}) = y_{dog}$
2. Confident $z^{(1)}_{dog-easy}[dog] > t$

URW = Uniform re-weighting



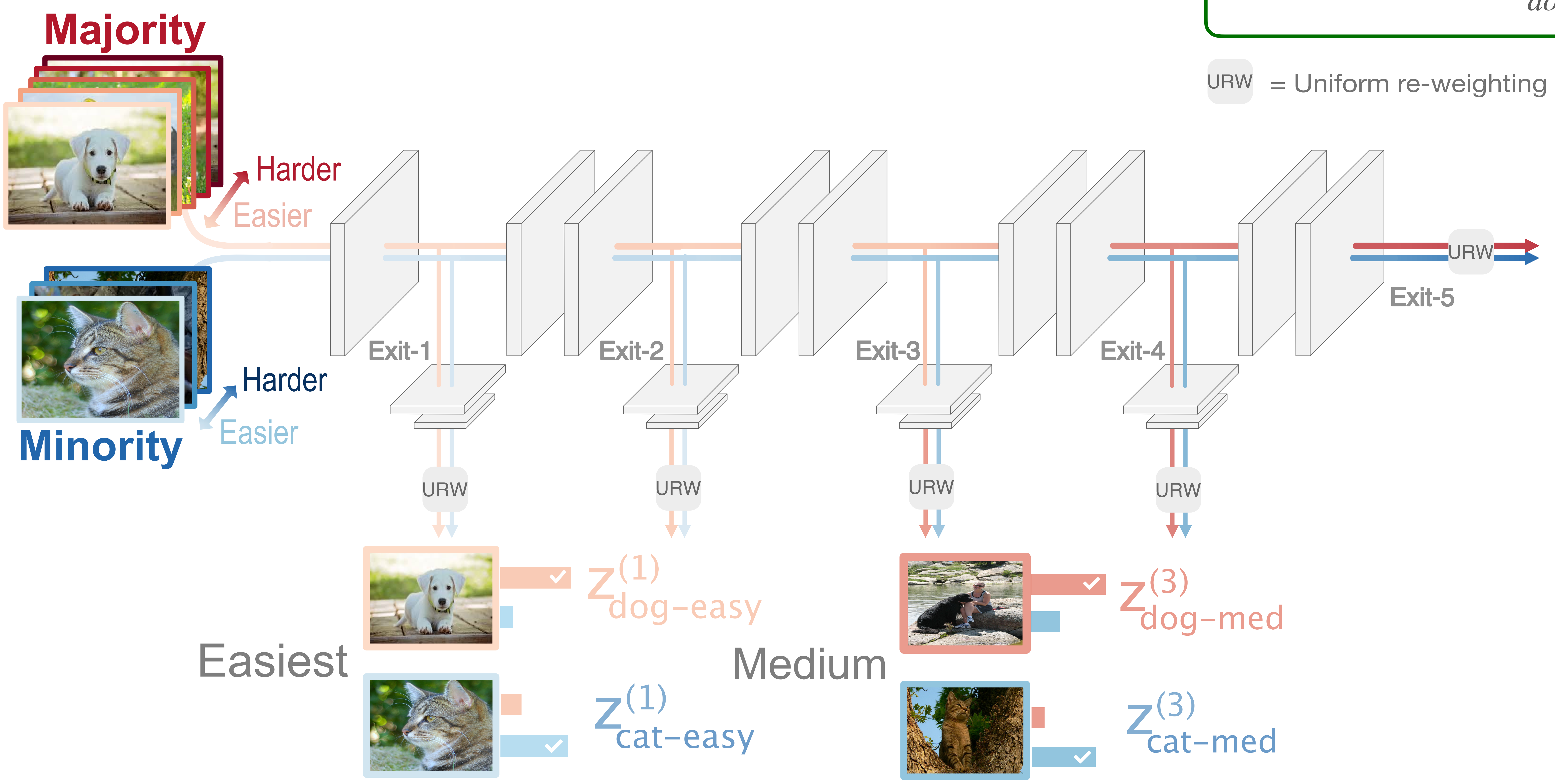
$L_{HAR} = L(z^{(1)}, y)$

HAR Framework

Early-Exit Conditions for Image

1. Correct $argmax(z^{(1)}_{dog-easy}) = y_{dog}$
2. Confident $z^{(1)}_{dog-easy}[dog] > t$

URW = Uniform re-weighting



$L_{HAR} = L(z^{(1)},y)$

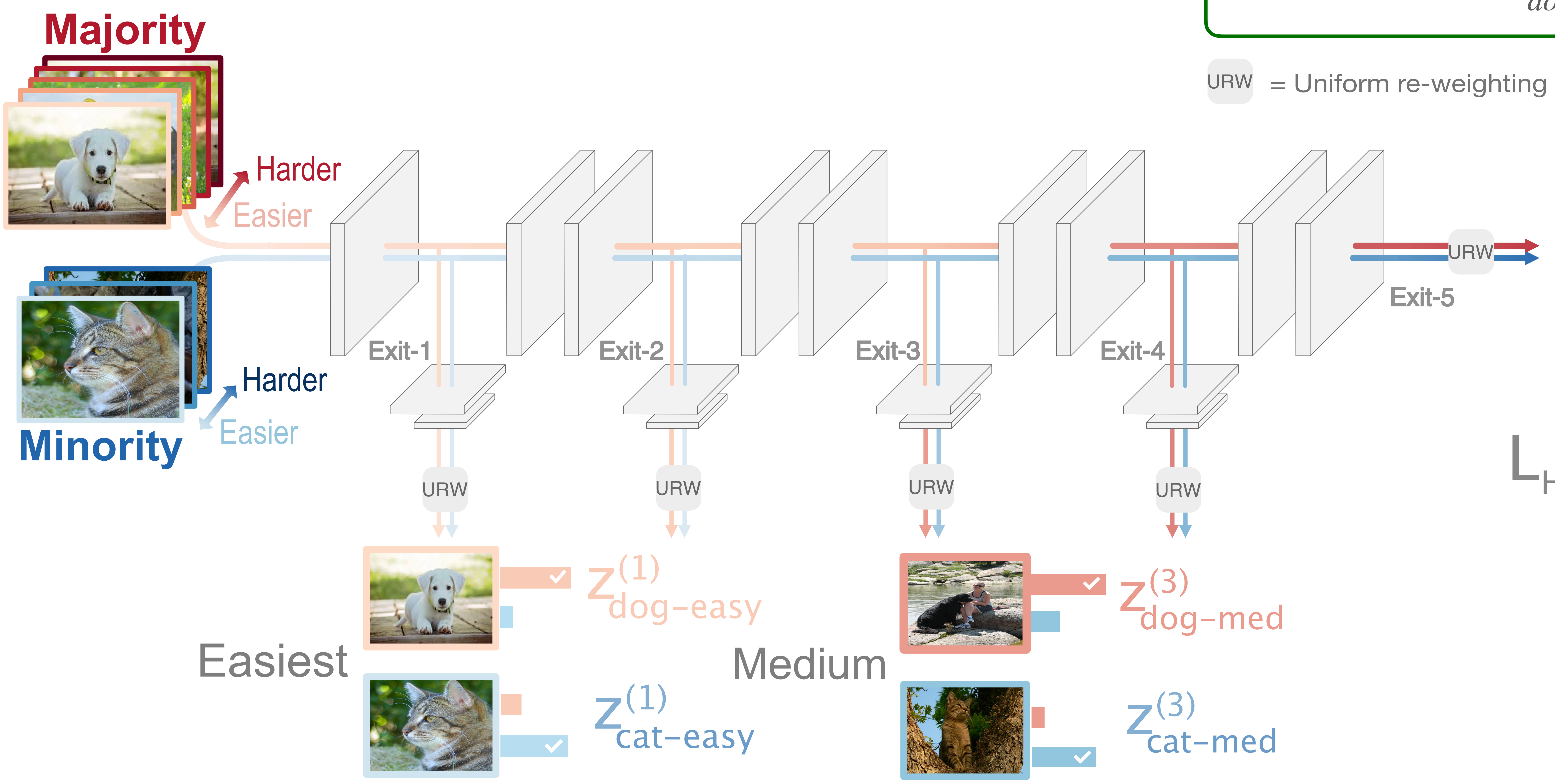
$L_{HAR} = L(z^{(1)},y) + L(z^{(2)},y) + L(z^{(3)},y)$

HAR Framework

Early-Exit Conditions for Image

1. Correct $argmax(z^{(1)}_{dog-easy}) = y_{dog}$
2. Confident $z^{(1)}_{dog-easy}[dog] > t$

URW = Uniform re-weighting



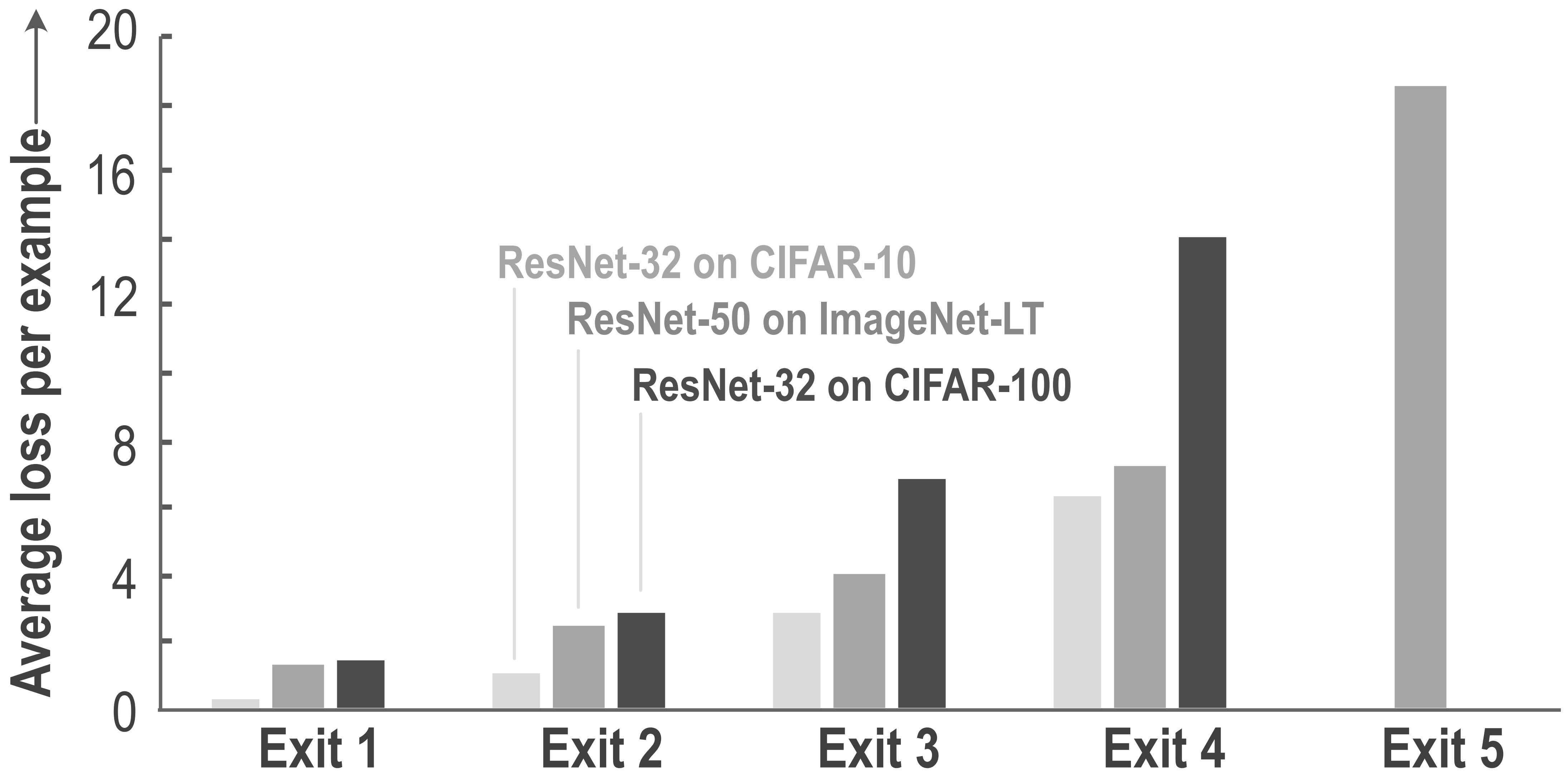
$$L_{HAR} = \sum_{i=1}^{k} L(z^{(i)}, y)$$

$$L_{HAR} = L(z^{(1)},y)$$

$$L_{HAR} = L(z^{(1)},y) + L(z^{(2)},y) + L(z^{(3)},y)$$

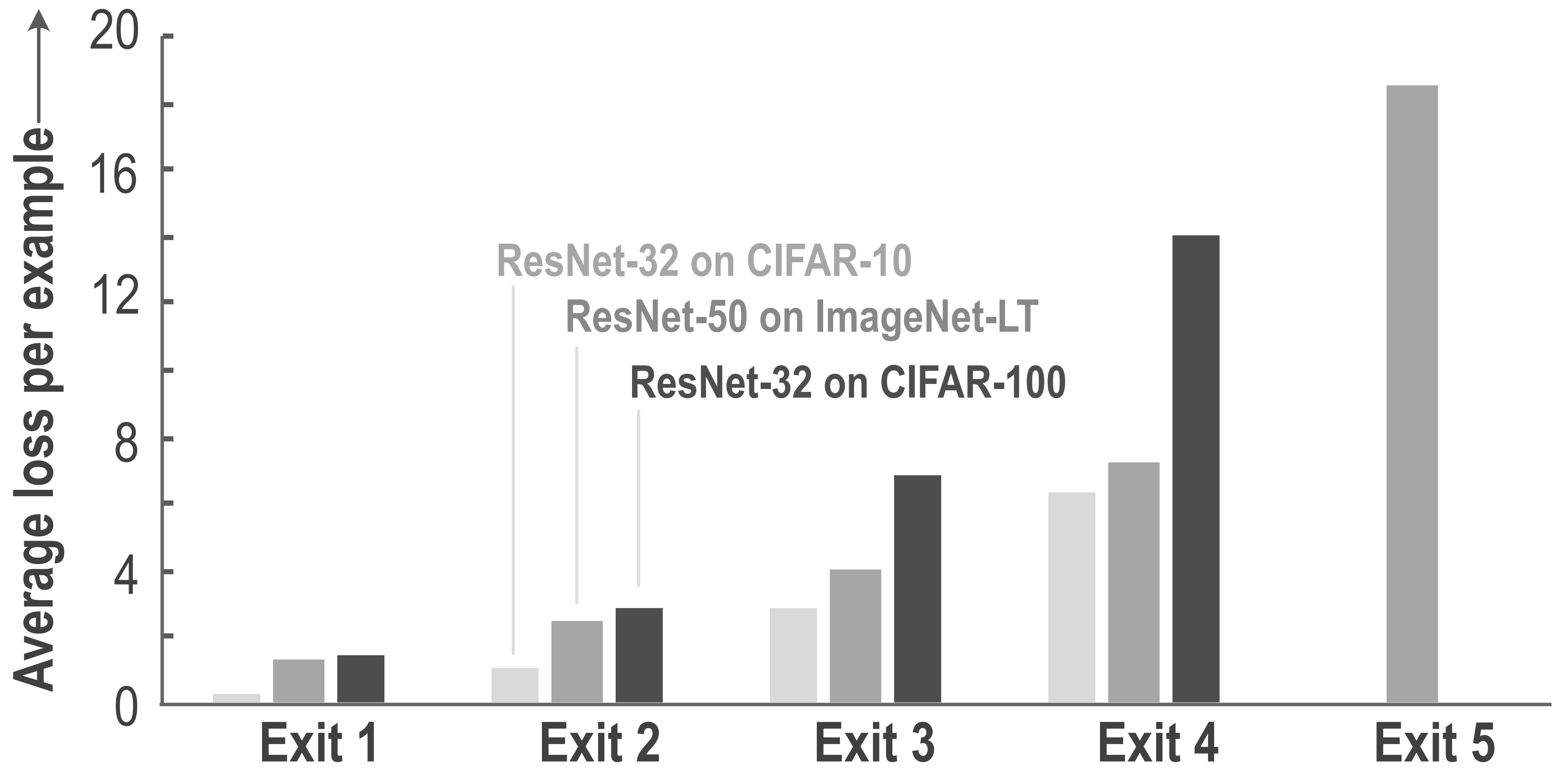
Results - 1

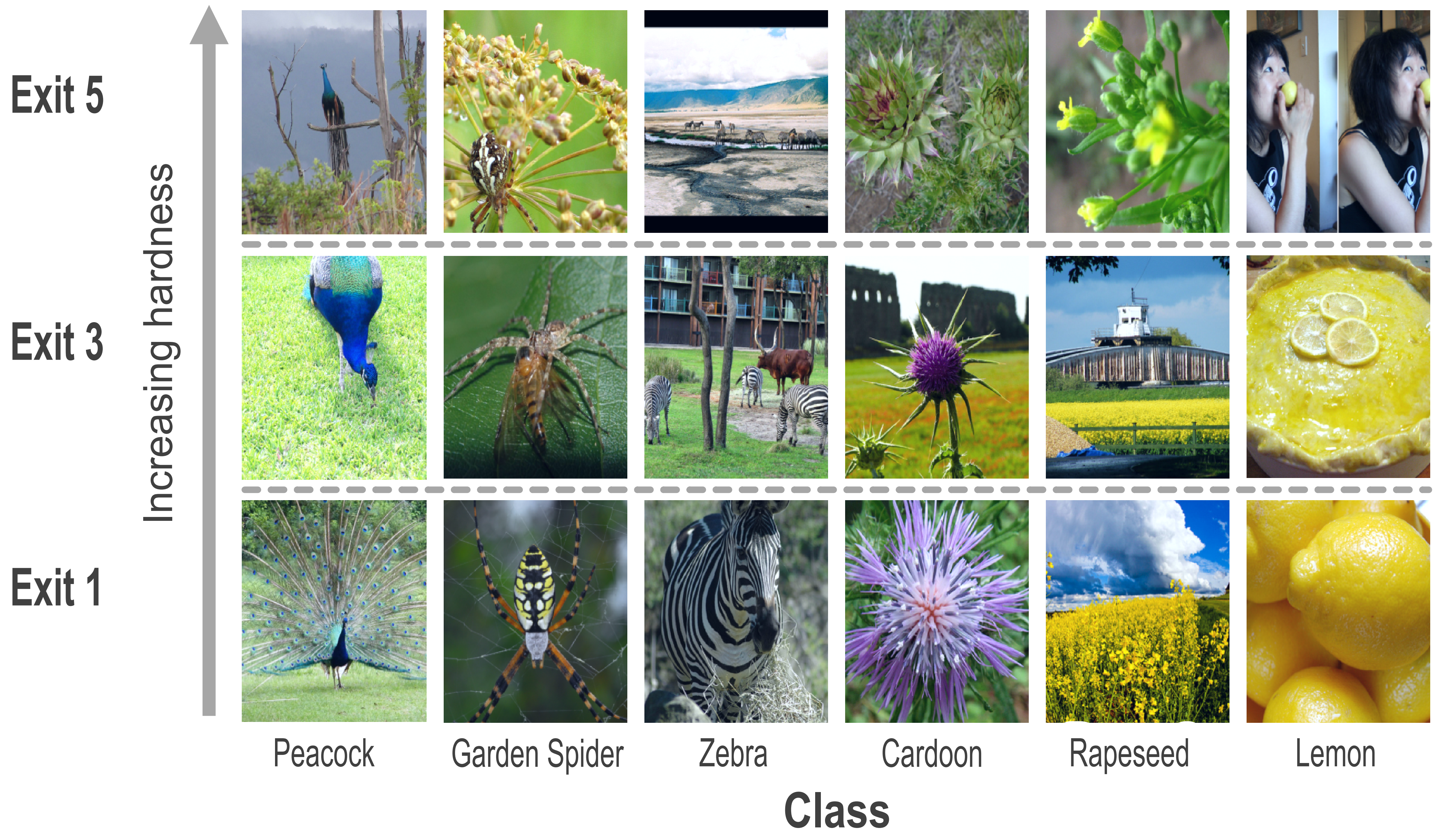
Outcome of early-exiting during training



Results - 1

Outcome of early-exiting during training





Results - 2

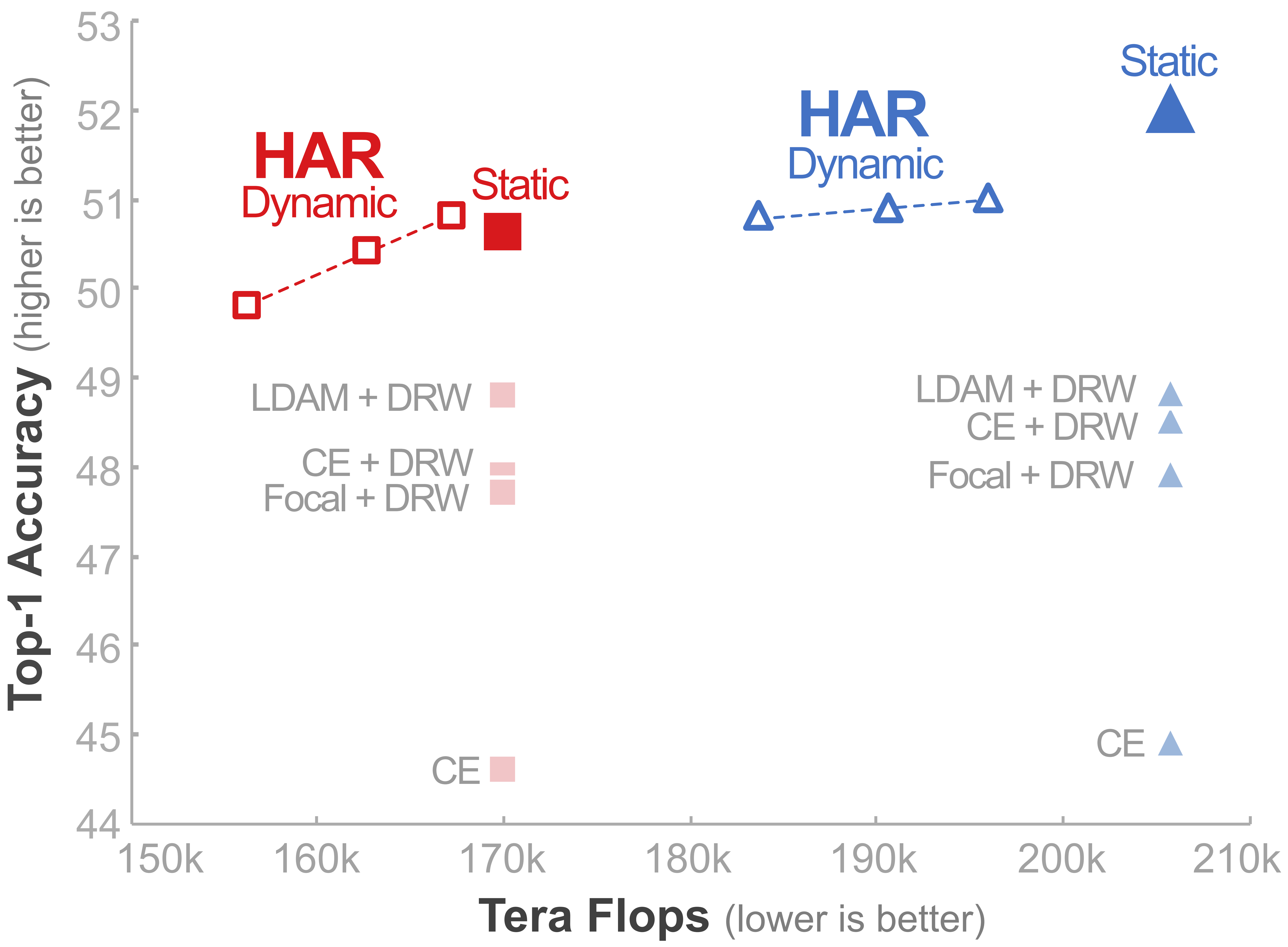
HAR's impact on accuracy

ResNet-50 on ImageNet-LT

Loss	Many	Med	Few	All
Focal ($\gamma = 0.5$)	63.5	38.5	13.6	44.7
Focal ($\gamma = 1.0$)	62.8	37.1	12.7	43.7
Focal ($\gamma = 2.0$)	62.1	36.6	11.9	43.1
CE	63.8	38.5	13.6	44.9
HAR (CE)	**63.9**	**39.9**	**16.0**	**45.9**
LDAM	64.9	39.1	12.6	45.5
HAR (LDAM)	**66.3**	**42.4**	**14.2**	**47.8**

Results - 3

HAR's impact on efficiency



Results - 3

HAR's impact on efficiency

	ImageNet LT				iNaturalist'18			
	Mny	Med	Few	All	Mny	Med	Few	All
CE†	63.8	38.5	13.6	44.6	72.7	63.8	58.7	62.7
CRT [6]††	58.8	44.0	26.1	47.3	69.0	66.0	63.2	65.2
LWS [6]††	57.1	45.2	29.3	47.7	65.0	66.3	65.5	65.9
τ-norm [6]††	56.6	44.2	27.4	46.7	65.6	65.3	65.9	65.6
Focal+DRW [21]†	59.5	44.6	27.0	47.9	66.1	66.0	64.3	65.4
CE+DRW [5]†	60.3	45.2	27.0	48.5	67.1	66.2	**65.4**	65.9
HAR(CE)+DRW (Our)	**60.7**	**45.5**	**27.7**	**48.9**	**67.4**	**66.3**	65.1	**66.0**
Flops Saving				21%				13%
LDAM + DRW [20]†	61.1	44.7	28.0	48.8	70.0	67.4	**66.1**	67.1
HAR(LDAM)+DRW (Our)	**63.8**	**47.2**	**28.1**	**51.0**	**72.2**	**69.0**	65.7	**68.0**
Flops Saving				5%				2%

Results - 3

HAR's impact on efficiency

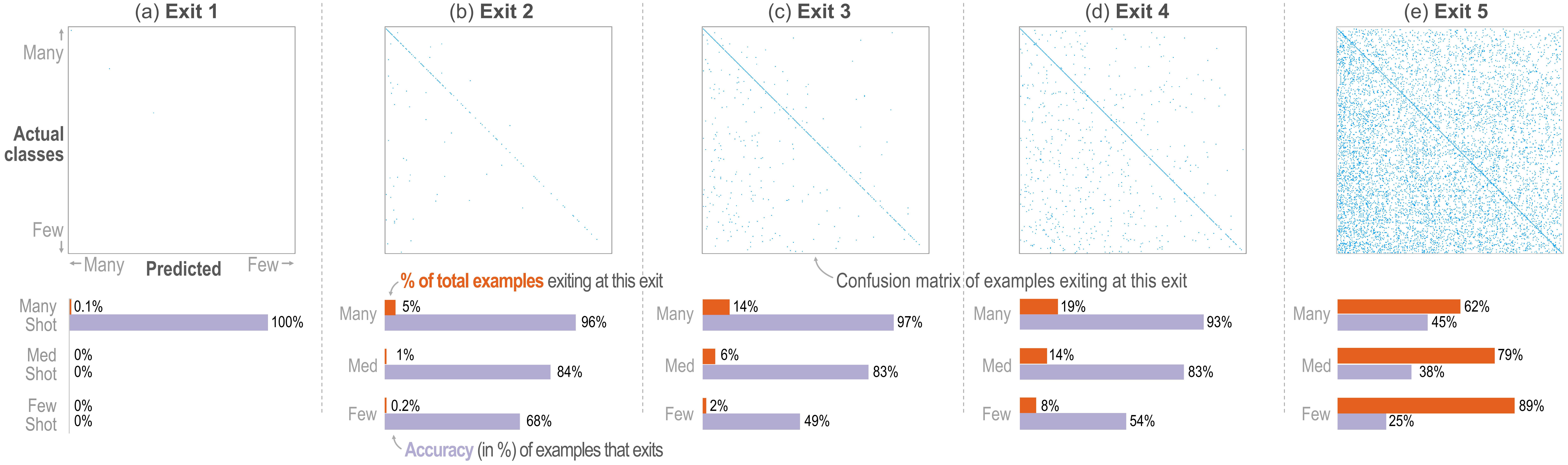
	ImageNet LT				iNaturalist'18			
	Mny	Med	Few	All	Mny	Med	Few	All
CE†	63.8	38.5	13.6	44.6	72.7	63.8	58.7	62.7
CRT [6]††	58.8	44.0	26.1	47.3	69.0	66.0	63.2	65.2
LWS [6]††	57.1	45.2	29.3	47.7	65.0	66.3	65.5	65.9
τ-norm [6]††	56.6	44.2	27.4	46.7	65.6	65.3	65.9	65.6
Focal+DRW [21]†	59.5	44.6	27.0	47.9	66.1	66.0	64.3	65.4
CE+DRW [5]†	60.3	45.2	27.0	48.5	67.1	66.2	**65.4**	65.9
HAR(CE)+DRW (Our)	**60.7**	**45.5**	**27.7**	**48.9**	**67.4**	**66.3**	65.1	**66.0**
Flops Saving				21%				13%
LDAM + DRW [20]†	61.1	44.7	28.0	48.8	70.0	67.4	**66.1**	67.1
HAR(LDAM)+DRW (Our)	**63.8**	**47.2**	**28.1**	**51.0**	**72.2**	**69.0**	65.7	**68.0**
Flops Saving				5%				2%

	CIFAR-10 LT			CIFAR-100 LT		
	100×	50×	10×	100×	50×	10×
CE†	70.4	74.8	83.6	28.3	43.9	55.7
Focal [21]††	70.4	76.7	86.7	28.4	44.3	55.8
Mixup [48]††	73.1	77.8	87.1	39.5	45.0	58.0
Manifold Mixup [49]††	73.0	78.0	87.0	38.3	43.1	56.5
CE+DRW [5]†	76.3	80.0	87.6	41.4	46.0	58.3
HAR(CE)+DRW (Our)	**76.8**	**80.8**	**87.6**	**42.5**	**47.1**	**58.7**
Flops saving	32%	29%	26%	11%	11%	10%
LDAM+DRW [20]†	77.0	81.4	87.6	42.0	46.6	58.7
HAR(LDAM)+DRW (Our)	**78.1**	**82.4**	**88.0**	**43.1**	**47.5**	**58.9**
Flops Saving	15%	21%	20%	0%	2%	3%

Results - 4

Analysis: Which examples exit early?

Predictions at earlier exits predominantly belong to many- and medium-shot classes, and are more accurate.



Analysis: Impact of each exit

ResNet-50 on ImageNet-LT

# early-exits	Config	Many	Med	Few	All
0	CBBBBE	61.1	44.7	28.0	48.8
1	CEBBBBE	62.3	44.9	25.9	49.0
	CBEBBBE	63.0	45.9	26.7	49.9
	CBBEBBE	63.3	46.6	27.8	50.5
	CBBBEBE	62.5	46.0	27.6	49.8
	Average	62.8	46.9	27.0	49.8
2	CEBEBBBE	63.0	46.0	27.6	50.0
	CEBBEBBE	63.4	46.6	27.4	50.5
	CEBBBEBE	62.2	45.6	26.8	49.4
	CBEBEBBE	63.4	46.3	27.5	50.3
	CBEBBEBE	63.1	46.9	27.9	50.6
	Average	62.9	46.2	27.4	50.0
3	CEBEBEBBE	63.5	46.7	27.8	50.6
	CEBEBBEBE	62.6	45.7	27.7	49.8
	CEBBEBEBE	63.5	47.0	28.1	50.8
	CBEBEBEBE	63.2	46.9	28.0	50.7
	Average	63.2	46.6	28.0	50.5
4	**CEBEBEBEBE**	**63.6**	**47.4**	**28.4**	**51.1**

Exits in the middle are more important

Accuracy improves with # early exits

Conclusion

Thank you!

Hardness Aware Reweighing (HAR) framework

- Uses hardness to unlock generalization performance

Benefits of HAR

- State-of-the-art accuracy
- Compute savings during inference
- Plug and play support for existing loss functions

Extensive evaluation & ablation on 4 datasets